An Investigation of the Knowledge Overlap between Pharmacogenomics and Disease Genetics[a]

Binglan Li[1], Michelle Whirl-Carrillo[1], Matt W. Wright[1], Larry Babb[2], Heidi L. Rehm[2,3,4,5], Teri E. Klein[1,6,+]

1 Department of Biomedical Data Science, Stanford University, 1265 Welch Road, Stanford, CA 94305
2 Broad Institute of MIT and Harvard, Cambridge, MA, 02142
3 Massachusetts General Hospital, Boston, MA, 02142
4 Brigham and Women's Hospital, Boston, MA, 02142
5 Harvard Medical School, Boston, MA, 02142
6 Department of Medicine (BMIR), Stanford University, 1265 Welch Road, Stanford, CA 94305
+ Corresponding author, email: teri.klein@stanford.edu

Precision medicine faces many challenges, including the gap of knowledge between disease genetics and pharmacogenomics (PGx). Disease genetics interprets the pathogenicity of genetic variants for diagnostic purposes, while PGx investigates the genetic influences on drug responses. Ideally, the quality of health care would be improved from the point of disease diagnosis to drug prescribing if PGx is integrated with disease genetics in clinical care. However, PGx genes or variants are usually not reported as a secondary finding even if they are included in a clinical genetic test for diagnostic purposes. This happens even though the detection of PGx variants can provide valuable drug prescribing recommendations. One underlying reason is the lack of systematic classification of the knowledge overlap between PGx and disease genetics. Here, we address this issue by analyzing gene and genetic variant annotations from multiple expert-curated knowledge databases, including PharmGKB, CPIC, ClinGen and ClinVar. We further classified genes based on the strength of evidence supporting a gene's pathogenic role or PGx effect as well as the level of clinical actionability of a gene. Twenty-six genes were found to have pathogenic variation associated with germline diseases as well as strong evidence for a PGx association. These genes were classified into three sub-categories based on the distinct connection between the gene's pathogenic role and PGx effect. Moreover, we have also found thirteen RYR1 genetic variants that were annotated as pathogenic and at the same time whose PGx effect was supported by a preponderance of evidence and given drug prescribing recommendations. Overall, we identified a nontrivial number of gene and genetic variant overlaps between disease genetics and PGx, which laid out a foundation for combining PGx and disease genetics to improve clinical care from disease diagnoses to drug prescribing and adherence.

Keywords: Medical Genetics; Pharmacogenomics; Disease Genetics; Genetic Annotation; Gap in Knowledge; Knowledge Overlap.

Introduction

Implementation of genomics-based precision medicine in routine clinical care faces many challenges including the gap in knowledge between disease genetics and pharmacogenomics (PGx). Disease genetics interpret genetic variants for their pathogenicity to improve genetic disorder diagnosis and management. PGx interprets the influence of genetic variation on drug response. Ideally, these two active clinical fronts of precision medicine will together enhance clinical care from disease diagnosis to drug prescribing. This ideal situation, however, is currently hindered by a gap in defining the relationship between these areas of genomic knowledge.

This knowledge gap makes it difficult to seamlessly integrate disease genetics and PGx into routine clinical care. A clinical genetic test may be ordered solely for the purpose of disease diagnosis or preemptive surveillance. It is unclear how many genes or variants in the returned results of clinical genetic tests that are ordered for diagnostic purposes are also of PGx significance for drug prescribing, in effect, maximizing efficacy and decreasing potential adverse drug reactions.

Currently, a cumulatively large number of gene and variant annotations for disease genetics and PGx are hosted in various knowledge databases. ClinGen[1] is a National Human Genome Research Institute (NHGRI) funded collaborative effort that provides expert-curated clinically interpreted genes and variants. Diverse curation activities of ClinGen include variant pathogenicity assertions, classifications of gene-disease validity, gene-disease clinical actionability reports, and dosage sensitivity. ClinVar[2,3], a National Center for Biotechnology Information (NCBI) publicly-accessible genomic data resource, hosts more than a million genetic variant classifications including expert variant pathogenicity curation efforts from ClinGen's expert panels. Collection and curation of PGx knowledge and drug prescribing recommendations, on the other hand, are actively maintained and updated by the Pharmacogenomics Knowledgebase (PharmGKB)[4] and the Clinical Pharmacogenetics Implementation Consortium (CPIC)[5,6].

The discussions of the knowledge overlap between PGx and disease genetics are scarce while the PGx community has gradually shifted greater attention and efforts towards the implementation of PGx based on an evidence-based strategy in clinics[7]. Most of the investigations and reviews focus on the PGx genes with well-known relationships to disease, such as the gene *CFTR* implicated in cystic fibrosis[8]. However, there is a lack of systematic investigation of the shared genes and variants between PGx and disease genetics across large-scale curation and knowledge resources.

Here, we carried out a systematic classification of the knowledge overlap between PGx and disease genetics by investigating genetic annotations from PharmGKB[4], CPIC[5,6], and ClinGen-approved expert panels[1,2,3]. We then evaluated the number of disease-related genes that turned out to be genes harboring PGx variants; and the number of disease-related genetic variants that had PGx effects. The estimations from this study provided a foundation, supported by evidence-based curations, for integrating disease genetics and PGx in future clinical care.

1. Materials and Methods

1.1. *PharmGKB clinical annotations and curated drug labels*

PharmGKB[4] clinical annotations were retrieved from the PharmGKB Downloads webpage (clinicalAnnotation.zip from https://www.pharmgkb.org/downloads). Clinical annotations

summarize all the PharmGKB annotations of published evidence for relationships between a specific genetic variant (annotated with pharmacogenes when appropriate) and a medication with an assigned score-based level of evidence ranging from 1 to 4 (1 meeting the highest criteria). As of July 12, 2021, PharmGKB has collected and curated 4,858 clinical annotations for 1,052 unique genes. We excluded 9 cancer genome pharmacogenes listed on the PharmGKB VIP web page (https://www.pharmgkb.org/vips) as these genes are typically not returned in the results from a patient's PGx test for germline variation. We retained only the highest level of evidence for a gene-drug pair or a variant-drug pair if more than one clinical annotation exists for the pair.

PharmGKB monitors the US Food and Drug Administration (FDA) Table of Pharmacogenomic Biomarkers in Drug Labeling and annotates the drug labels from the list with PGx information. The full FDA biomarker list is available online (https://www.fda.gov/drugs/science-and-research-drugs/table-pharmacogenomic-biomarkers-drug-labeling). However, FDA's list of PGx biomarkers includes genes regardless of clinical impact on drug response. PharmGKB created a PGx Level system for annotating labels with the type of PGx information found on them and highlighting specific prescribing information where applicable. We retrieved the drugLabels.tsv annotation file from the PharmGKB Downloads page (drugLabels.zip from https://www.pharmgkb.org/downloads).

Dated July 12th, 2021, PharmGKB has curated 371 drug labels approved by the US FDA. We only retained drug label annotations that met the following criteria sequentially: (1) The annotated gene was currently on the FDA PGx biomarker list; (2) The gene was not one of the 9 cancer genome pharmacogenes; and (3) It had the PharmGKB-curated "Prescribing Info" tag. These criteria reduced the numbers from 371 FDA-approved drug label annotations down to 150 annotations.

1.2. *CPIC gene-drug pairs*

CPIC[5,6] levels of gene-drug pairs were downloaded from the CPIC website (https://cpicpgx.org/genes-drugs/) on July 12th, 2021. CPIC[5,6] has assigned levels to 442 gene-drug pairs as of July 30, 2021, encompassing 269 unique medications and 118 unique genes. Only pairs that have CPIC guidelines have had sufficient in-depth review of the evidence to be given a definitive level assignment. The remaining pairs have been given a provisional level assignment. We retained only the highest CPIC levels assigned to a gene if it appeared in multiple CPIC gene-drug pairs.

1.3. *ClinGen and ClinVar Annotation Resources*

ClinGen[1] provides various assertions on genes and variants. This study focused on the variant-level pathogenicity classifications by ClinGen's Variant Curation Expert Panels (VCEPs)[1], gene-disease validity assertions by the ClinGen's Gene Curation Expert Panels (GCEPs)[9] and clinical actionability reports by the Actionability Working Group (AWG)[10].

ClinGen variant pathogenicity annotations abide by the 5-tier pathogenicity nomenclature jointly developed by the American College of Medical Genetics and Genomics (ACMG) and the Association for Molecular Pathology (AMP)[11]. A full table of variant pathogenicity that was curated and approved by VCEPs was downloaded as a tab delimited file from the web page (https://erepo.clinicalgenome.org/evrepo/) on July 12th, 2021. rsIDs of the locations of these variants were retrieved from the ClinGen Allele Registry through the ClinGen API using the command "curl http://reg.clinicalgenome.org/alleles?file={id} --data-binary @<file of ClinGen allele registry

IDs>." Dated July 12th, 2021, ClinGen VCEPs have classified 2,423 Mendelian disorder-associated genetic variants, encompassing 40 unique genes and 41 unique diseases.

Gene-disease validity relates to the level of evidence to support a gene's role in disease[9] and is curated by ClinGen GCEPs. The table of curated clinical validity classification of gene-disease pairs was downloaded from ClinGen (https://search.clinicalgenome.org/kb/downloads) on July 12th, 2021. As of July 12th, 2021, 1,393 gene-disease pairs had been evaluated for clinical validity, encompassing 1,127 unique genes and 680 unique diseases.

For clinical validity classifications, we first retained the gene-disease pair with the strongest levels of evidence for a gene if a gene is associated with multiple diseases (Definitive > Strong > Moderate > Limited > No Known Disease > Disputed > Refuted). We removed the gene-disease pairs with "Limited" classification and below, which included "Limited" (limited evidence to support a causal role), "No Known Disease Relationship" (referring to genes for which the disease claim has no human evidence), "Disputed" (conflicting evidence questioning a causal role) and "Refuted" (convincing evidence refuting a causal role). Gene-disease pairs in these classification categories were removed due to a high prevalence of false positives or a definite assertion of the non-pathogenic roles of genes in these categories. To be specific, genes with "Definitive", "Strong", or "Moderate" clinical validity classification were retained in the analyses; and a gene with "Limited", "No Known Disease Relationship", "Disputed" or "Refuted" classification would persist if the gene was associated with another disease with "Definitive", "Strong", "Moderate" clinical validity classifications. 244 unique genes were removed which brought the number down to 883 unique genes with clinical validity annotations of "Moderate" or above.

The ClinGen AWG assessed the clinical actionability of a smaller number of gene-disease pairs using semiquantitative metric (SQM) scores, an evidence-based quantitative profile based on disease severity, likelihood of disease, effectiveness of interventions, nature of interventions (risks *vs* benefits), and level of evidence. We retrieved the "summary of overall scores" of the clinical actionability downloads for both adults and pediatrics from the ClinGen public-facing web-based portal (https://search.clinicalgenome.org/kb/downloads) on July 12th, 2021. Dated the time of download, ClinGen AWG had generated 594 clinical actionability reports (388 for adults and 206 for pediatrics). A gene-disease pair may have more than one SQM score due to assessments on multiple distinct interventions. Overall, the reports evaluated clinical actionability of 237 unique gene-disease pairs on variable interventions (212 unique genes and 121 unique diseases).

ClinVar[2,3] records were obtained from efforts at the Broad Institute on July 7th, 2021. The ClinVar data is managed through an automated real-time process that digests and normalizes the full weekly ClinVar variation archive XML files into a Google Cloud Big Query relational database preserving all ClinVar derived and submitted content. Each weekly ClinVar release is processed to contextualize every submitted assertion (SCV) record into one of four contexts: Germline Disease (has P/LP/VUS/LB/B/Risk classification), Somatic Cancer (has both a 'somatic' allele origin and in a cancer gene), Pharmacogenomic (has assertion of Drug Response) or Other. SCVs that had no assertion criteria provided (0-star SCVs) were marked excluded and did not contribute to the re-aggregation process. Only the SCV records and associated variant-level re-aggregated classifications and review statuses were used in the analysis. The variant aggregate classification (VCV) and review status was then re-computed for any context category that had one or more SCVs for a given variant. Only VCVs for Germline Disease were used in the analysis.

1.4. *Data Analysis*

Data analysis and visualization were performed using *dplyr*, *ggplot2*, and *nVennR* in R v4.0.4.

1.5. *Classification of Genes*

A gene with strong PGx support was defined as a gene that had a preponderance of evidence supporting its effect on drug response as found in PharmGKB clinical annotations or, as a gene that had available or pending expert-curated CPIC guidelines. Technically, a gene with strong PGx support would satisfy one of the following criteria:

(1) A PharmGKB curated clinical annotation with a level of evidence 1 or 2; or,

(2) A CPIC level A or B gene-drug pairs; or,

(3) Included in the 150 PharmGKB curated FDA-approved drug labels described in **Methods 2.1**.

A gene with sparse PGx support is a gene with limited evidence supporting its effect on drug responses as found in PharmGKB clinical annotations; and/or, has not had an in-depth evidence review by CPIC or had been reviewed by CPIC and found to have inadequate evidence for a prescribing recommendation. A gene with sparse PGx support would satisfy one or a combination of the following criteria:

(1) A PharmGKB curated clinical annotation at levels of evidence 3 or 4 only; and/or

(2) A CPIC provisional level B/C gene-drug pairs or levels C or D gene-drug pairs.

Disease genetics-wise, a gene was considered to have strong evidence supporting its causal role in a certain gene-disease relationship or clinical actionability. A disease gene with strong clinical support meets any one of the following criteria:

(1) A ClinGen curated (likely) pathogenic variant; or

(2) A "Definitive" or "Strong" ClinGen curated clinical validity classification for a gene-disease relationship in human; or

(3) An actionability report with an integrative, evidence-based semiquantitative metric (SQM) score 8 for "medical actionability" and, at the same time, has a likelihood of disease of levels of evidence A or B or an efficacy of intervention of levels of evidence A or B. In total, 142 genes satisfied this specific requirement where 131 of them had "Definitive" or "Strong" clinical validity classifications and 11 had no clinical validity classifications yet (annotations are likely pending).

Genes that did not meet any of the abovementioned criteria were considered to have only sparse evidence supporting its causal role in a certain gene-disease relationship and to have a lack of clinical actionability. A disease gene with sparse clinical support would satisfy one of the following criteria:

(1) Only had ClinGen curated (likely) benign variants (only *KLLN*); and/or

(2) "Moderate" ClinGen curated clinical validity classification for a gene-disease relationship in humans; and/or

(3) No actionability report with an integrative, evidence-based semiquantitative metric (SQM) score 8 for "medical actionability"; if the SQM score is 8, has a likelihood of disease of only levels of evidence C or D and an efficacy of intervention of only levels of evidence C or D. Genes with strong pathogenic support, such as those with "Definitive" or "Strong" clinical validity classifications, would not fall into this category. This specific requirement identified two genes with "Moderate" clinical validity classification, and eighteen genes with no clinical validity classifications yet (annotations are likely pending).

To clarify, the lack of clinical actionability does not indicate pathogenicity of a gene in a genetic disease. Most genes implicated in disease are not clinically actionable even though they have definitive evidence suggesting their pathogenic role in a disease. In this study, disease genes with sparse clinical support can become a gene with strong pathogenic evidence over time but currently lack a preponderance of evidence to be asserted as the rest of the "Definitive" or "Strong" gene-disease pairs. Disease genes with sparse clinical support should not be mistaken as non-pathogenic genes indefinitely.

Secondly, ClinGen does not curate variant pathogenicity if the gene is not validly associated with a genetic disease. Genes that had variant pathogenicity annotations all fell into the category of disease genes with strong clinical support, except *KLLN*. *KLLN* variants were in the promoter region of *PTEN* and were assigned to *KLLN* in this case due to the shared promoter between *KLLN* and *PTEN*. In addition, the *KLLN* variants were annotated as uncertain significance or (likely) benign variants. Given these reasons, we removed the gene *KLLN* from the ClinGen variant pathogenicity annotations and prevented further analysis of the gene based on variant pathogenicity annotations.

2. Results

2.1. *Summary*

Dated July 12th, 2021, we retrieved and investigated disease genetic annotations and PGx annotations for both genes and genetic variants from PharmGKB[4] (clinical annotations and expert-curated drug labels), CPIC[5,6] (gene-drug pairs), ClinGen[1] (variant pathogenicity, gene-disease clinical validity, clinical actionability reports), and ClinVar[2,3] (aggregate ClinVar assertions) (**Table 1**).

Table 1. Summary statistics of raw gene and variant annotations retrieved from PharmGKB, CPIC, and ClinGen, dated July 12th, 2021.

Data Sources	Description	Total Annotations	Unique Genes or Variants	PMID
CPIC	Gene-Drug Pairs	442	118 Genes	21270786
PharmGKB	Clinical Annotations	4,858	3,126 Variants; 1,052 Genes	22992668
	Drug Labels	799 (371 FDA approved)	159 Genes	
ClinGen	Variant Pathogenicity	2,423	2407 Variants; 40 Genes	26014595
	Clinical Validity	1,393	1127 Genes	
	Actionability Reports	594	237 Genes	
	ClinVar*	1,036,863		Pending

* The data was last updated and aggregated from ClinVar on July 7th, 2021. Statistics are subject to updates in the coming release.

Overall, disease genetics-wise, there were 914 unique genes with expert-reviewed variant pathogenicity, gene-disease validity, or actionability reports (**Figure 1**). We classified genes based on the evidence supporting the pathogenic role or clinical actionability of a gene. 773 (85%) of the 915 genes have strong support of the gene's pathogenic role or clinical actionability in a certain disease. 141 (15%) of the 914 genes have sparse evidence supporting the pathogenic role or clinical actionability of the gene in a certain disease. 914,253 variant classifications associated with germline disease were retained from the ClinVar data based on ClinGen's re-aggregation of ClinVar VCVs.

A total of 1,078 unique genes were recorded or curated in one or more of the three investigated PGx data sources (**Figure 1**). We also classified genes based on the strength of evidence supporting

the PGx effect or drug prescribing actionability of a gene. Eighty-eight (8%) of the 1,078 unique genes had strong PGx support for their effect or actionability. The other 990 (92%) curated genes had sparse evidence supporting their PGx effects or clinical actionability.

2.2. *Overlap between genes with strong PGx support and disease genes with strong clinical support*

We first explored the knowledge overlaps between disease genetics and PGx by investigating the number of common genes that have strong supporting evidence in both knowledge domains. In total, 26 genes had strong evidence supporting their pathogenic role in a disease or PGx effect (**Figure 1**). These 26 genes could be classified into three sub-categories based on the connection between their pathogenic role and PGx effect (**Table 2**).

		PGx	
		Genes with strong support (N = 88)	Genes with sparse support (N = 990)
Disease Genetics	Genes with strong support (N = 773)	26	86
	Genes with sparse support (N = 141)	1	12
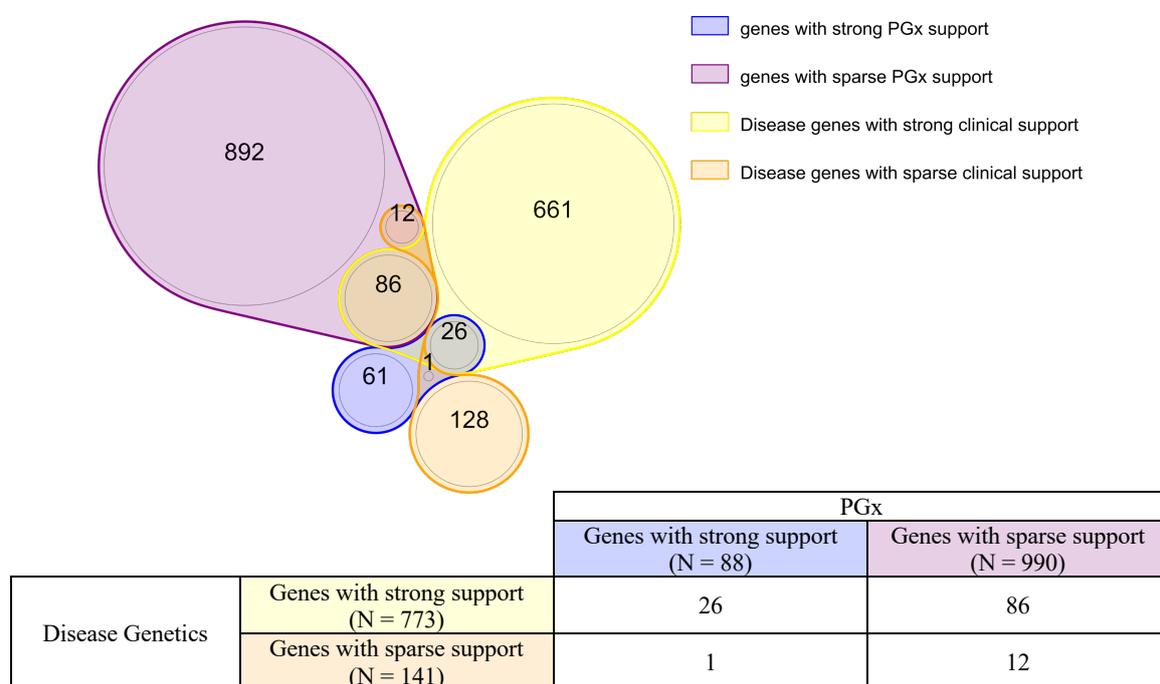
Figure 1. Overlap between PGx and disease genetics at the level of gene based on annotations or curations from PharmGKB, CPIC, and ClinGen and ClinVar. For either PGx or disease genetics, genes were dichotomously classified into genes with strong or sparse support based on level of supporting evidence. The **Venn diagram** showed the overlaps between genes with strong PGx support (**blue**), genes with sparse PGx support (**purple**), disease genes with strong clinical support (**yellow**), and disease genes with sparse clinical support (**orange**). The numbers in the Venn diagram denoted the number of genes in each class, either exclusive genes in a class or intersections with other classes of genes. An alternative **table** was provided with summary statistics of the total numbers of genes in each class, as well as the numbers of overlapping genes between classes.

The first class included 2 genes (*RYR1* and *CACNA1S*) whose pathogenic role was directly related to their PGx effects (**Table 2**). Carriers of specific pathogenic genotypes can live without a clinical phenotype unless exposed to certain drugs, *e.g.*, *RYR1* and *CACNA1S* for malignant hyperthermia (MH) upon potent volatile anesthetic agents[12]. Carriers of specific variants in the *RYR1* and *CACNA1S* genes are at increased risk of MH, in comparison to noncarriers, when administered potent inhalational anesthetics such as desflurane[12].

The second class encompassed 12 genes. Due to, in the worst case, lethal adverse drug reactions, certain medications are contraindicated in individuals with specific genetic disorders or carriers of

specific genotypes in these genes. For example, pathogenic variants in the *OTC* gene can lead to ornithine carbamoyltransferase deficiency (MONDO:0010703), a type of urea cycle disorders (UCDs)[13]. As valproic acid can suppress the urea cycle, this medication is contraindicated in patients with UCDs to prevent potentially fatal side effects[14].

The third class covered 16 genes. Contrary to the second class, medications were indicated for individuals with certain genetic conditions caused by the pathogenic genetic variations in the genes. For example, velaglucerase alfa is a long-term therapeutic medication for type 1 Gaucher disease patients who are deficient of *GBA* enzyme activity[15].

Some genes were classified to more than one category due to interactions with different drugs and drug substrate-specific effects. For instance, valproic acid is contraindicated in patients with UCDs caused by pathogenic *POLG* or *OTC* variants, while sodium benzoate/sodium phenylacetate and sodium phenylbutyrate are indicated treatments for certain subtypes of the UCDs.

Table 2. Three classes of genes that had both strong pharmacogenetic support and strong pathogenic support

Class	Relationship with Drugs	Genes	Drugs
I	Contraindication	*CACNA1S, RYR1*	Inhalational anesthetics
II	Contraindication	*ALDH5A1*	gamma hydroxybutyric acid/sodium oxybate
		POLG, CPS1, OTC	valproic acid
		F5	avatrombopag, eltrombopag and lusutrombopag
		F5, MTHFR, PROC, PROS1, SERPINC1	ethinyl estradiol / norelgestromin
		HPRT1	mycophenolic acid
		MTHFR	methotrexate
		POLG	divalproex sodium
		PROC, PROS1, SERPINC1	synthetic conjugated estrogens, estradiol / progesterone
		PROC, PROS1	warfarin
III	Indication	*ASL, ASS1, CPS1, NAGS, OTC*	sodium benzoate/sodium phenylacetate
		ASS1, CPS1, OTC	sodium phenylbutyrate
		BRCA1, BRCA2	talazoparib
		CHRNA4	eculizumab
		DMD	viltolarsen/golodirsen/eteplirsen
		GAA	alglucosidase alfa
		GALNS	elosulfase alfa
		GBA	velaglucerase alfa
		GLA	migalastat
		LMNA	lonafarnib
		NAGS	carglumic acid
		SCN1A	carbamazepine
		TPP1	cerliponase alfa

2.3. *Genes with strong support in one field but not the other*

We were also interested in finding genes associated with PGx but not germline disease and vice versa (**Figure 1**). Gene annotations may be absent in PGx or disease genetics, causing a gene to have sparse supporting evidence in this study, due to the lack of precedent studies and experimental

evidence (despite the potential functional impact of the gene) or insufficient time and/or resources for curating either PGx or disease-gene associations. It may be worth prioritizing research and curations for these genes given the knowledge about the gene in another research field.

VKORC1 was a disease gene that had sparse clinical support, but the same time, a gene with strong PGx support. Meanwhile, 86 genes that had sparse PGx support turned out to be disease genes that had strong clinical support per ClinGen annotations (**Figure 1**). Six of these genes are part of the extended ADME genes (absorption, distribution, metabolism, and excretion)[16], including *ATP7A, ATP7B, FMO3, HNF4A, SLC22A5,* and *KCNJ11*. It should be noted that ClinGen had only curated gene-disease validity on 1,127 genes at the time of this analysis, despite there being 4,518 genes in OMIM with at least one phenotype associated and therefore this number is assumed to be an underestimate.
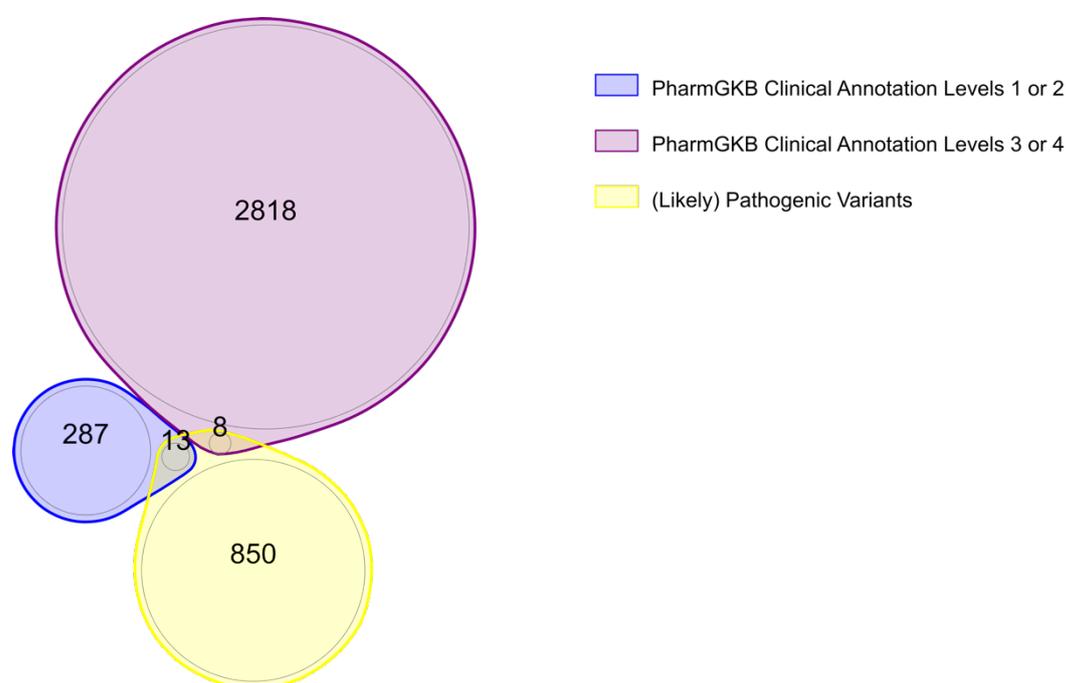
Figure 2. Overlapping variants between PharmGKB clinical annotations and ClinGen variant pathogenicity annotations. **Blue** area represented variants with PharmGKB clinical annotations of levels of evidence 1 or 2 that met a moderate to high criterion of level of evidence supporting the variant-drug relationship during the process of PharmGKB curation. **Purple** area followed a similar logic but represented variants with PharmGKB clinical annotations of levels of evidence 3 or 4 that had a low level of PGx evidence. **Yellow** area represented pathogenic or likely pathogenic variants curated by ClinGen's variant pathogenicity curation effort as of July 12th, 2021. We found in total 13 genetic variants shared between PharmGKB clinical annotation levels 1 or 2 and ClinGen's curated (likely) pathogenic variants.

2.4. *Overlapped pathogenic variants and PGx variants*

ClinGen Variant Pathogenicity curation activity shared 23 genetic variants with PharmGKB clinical annotations. The low number of overlapping genetic variants was not surprising as ClinGen variant curation efforts are only approved for 40 genes and there were only 9 shared genes between the two annotation sources to begin with (*TP53, RYR1, RAF1, PAH, LDLR, PTEN, CDH23, POLG,* and *ITGB3*). Thirteen overlapping variants were in *RYR1*, eight in *PAH*, and one each in *ITGB3* and *TP53* (**Supplementary Table**). **Figure 2** showed the overlap between PGx and (likely) pathogenic

variants curated by ClinGen's variant pathogenicity curation effort as of July 12th, 2021. Interestingly, 2 *RYR1* variants (rs144336148, NC_000019.10:g.38455472G>A; and rs111888148, NC_000019.10:g.38455463G>A) did not overlap. They were listed by ClinGen's Malignant Hyperthermia Susceptibility Variant Curation Expert Panel as uncertain significance[18] but were considered "Malignant Hyperthermia associated" by the CPIC guideline[12]. Due to this discovery, these variants are now under re-evaluation by CPIC based on the more recent ClinGen publication[18].

We expanded the search of overlapping genetic variants by comparing PharmGKB clinical annotations with the entire ClinVar database, using the re-aggregated VCV-level classifications calculated by ClinGen. We found 481 overlapped genetic variants between the germline disease-related ClinVar records and PharmGKB clinical annotations. Of the 481 overlapped genetic variants, 106 had PharmGKB clinical annotations at level 1 or 2 (**Supplementary Table**). Of these 106 genetic variants, 56 were classified as pathogenic or risk factors for germline diseases. The residual 50 genetic variants were classified as benign, uncertain significance, or conflicting interpretations. On the other hand, there were 46 of the 481 overlapped genetic variants that had PharmGKB level 3 clinical annotations (on genes *BCHE, CFTR, GLA, KCNH2, PAH, SCN4A,* and *UGT1A9*) and were also annotated as (likely) pathogenic or risk factor variants.

3. Discussion

We evaluated the knowledge overlap between disease genetics and PGx to understand the practicality and benefits of integrating the two knowledge domains in clinical care. We analyzed thousands of PGx and pathogenic annotations for genes and variants from several large publicly available knowledge databases, including PharmGKB[4], CPIC[5,6], ClinGen[1] and ClinVar[2,3]. Of the 914 annotated disease-associated genes, 773 (84%) had strong support of the gene's pathogenic role or clinical actionability in a certain disease. Of the 1,078 annotated PGx genes, 88 (8%) had strong support for their clinical actionability or PGx effect. Only 26 genes were shared between these 2 gene categories, indicating the knowledge overlaps between disease genetics and PGx (**Figure 1**).

The 26 genes were classified into three sub-categories based on the intra-connection of their pathogenic role and PGx effect (**Table 2**). Given specific genetic variants, medications were indicated or contraindicated in variant carriers in a drug-specific way. Genes implicated in urea cycle disorders (UCDs) (*POLG, CPS1,* and *OTC*) belonged to two classes that had opposite, drug-specific prescribing indications. Some drugs are used to treat the UCDs, while some will exacerbate the conditions in UCD patients and are contraindicated. It will be interesting to learn whether there are more cases of drug-specific prescribing recommendations for patients with genetic diseases.

Deficiency of studies or data in non-European or non-Eurocentric populations may contribute to the sparse PGx evidence support for the genes that we classified in this study. *SLC22A5* causes systemic primary carnitine deficiency disease (MONDO:0008919)[19] and is a transporter gene categorized by Hovelson *et al.*[16]. Two PGx studies investigated the effect of this gene (*SLC22A5*, aka *OCTN2*) in Asian populations (Singapore Chinese and Indian)[20,21]. The studies found that certain *SLC22A5* variant carriers were more susceptible to inhibition by the drugs cimetidine, pyrilamine and verapamil than those with no variants[20,21]. However, no follow-up reports were found.

PGx and disease genetics can each benefit by sharing the knowledge from the other domain. Combining PGx curations with disease genetics curations will provide a more comprehensive understanding of genetic conditions and guide clinical decision making. For example, the mitochondrial gene *MT-RNR1* is not only the causal gene for congenital or late-onset hearing

loss[22,23], but also a drug risk biomarker for aminoglycosides-induced hearing loss[24]. Patients with *MT-RNR1*-driving late-onset hearing loss will benefit from clinical genetic tests that note whether there is a known association with a PGx effect when returning pathogenic variants to avoid aminoglycosides that may worsen their hearing loss or lead to hearing loss in family members exposed to the drug.

The findings in this study will be updated over time as discoveries and evidence become available enhancing our understanding of PGx and disease genomics. For example, "Moderate" clinical validity classifications for gene-disease pairs may become "Definitive" due to growing evidence over time. Nonetheless, knowledge curation takes a nontrivial amount of time and resources but is critical for the translation of basic science to the bedside. Sometimes, the gaps in knowledge may be due to absence of available published data when the information is being curated.

PGx testing is becoming more common place in clinical care and may be covered by Medicare/Medicaid and commercial insurance[25]. Integration of disease genetics and PGx will benefit patients from the point of disease diagnosis to drug prescribing and adherence.

4. Supplementary data

Supplementary document is available at https://docs.google.com/document/d/1I6Ga8cZvuOH1HUqH-MsIPpeWsVMVNfF0hb-lP1URnmA/edit?usp=sharing.
Supplementary table is available at https://docs.google.com/spreadsheets/d/136YJMmx-Q6fNYza2gyYFwziBxhMkfOPXjloiQo6sZEA/edit?usp=sharing.

5. Acknowledgments

6. References

1. Rehm, H. L. *et al.* ClinGen--the Clinical Genome Resource. *N Engl J Med* **372**, 2235–42 (2015).
2. Landrum, M. J. *et al.* ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res* **42**, D980-5 (2014).
3. Landrum, M. J. *et al.* ClinVar: improvements to accessing data. *Nucleic Acids Res* **48**, D835–D844 (2020).
4. Whirl-Carrillo, M. *et al.* Pharmacogenomics knowledge for personalized medicine. *Clin Pharmacol Ther* **92**, 414–7 (2012).
5. Caudle, K. *et al.* Incorporation of Pharmacogenomics into Routine Clinical Practice: the Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline Development Process. *Curr. Drug Metab.* **15**, 209–217 (2014).
6. Relling, M. V. & Klein, T. E. CPIC: Clinical Pharmacogenetics Implementation Consortium of the Pharmacogenomics Research Network. *Clin Pharmacol Ther* **89**, 464–7 (2011).
7. Relling, M. V. & Evans, W. E. Pharmacogenomics in the clinic. *Nature* **526**, 343–50 (2015).
8. Clancy, J. P. *et al.* Clinical Pharmacogenetics Implementation Consortium (CPIC) Guidelines for Ivacaftor Therapy in the Context of CFTR Genotype. *Clin. Pharmacol. Ther.* **95**, 592–597 (2014).

9. Strande, N. T. *et al.* Evaluating the Clinical Validity of Gene-Disease Associations: An Evidence-Based Framework Developed by the Clinical Genome Resource. *Am J Hum Genet* **100**, 895–906 (2017).

10. Hunter, J. E. *et al.* A standardized, evidence-based protocol to assess clinical actionability of genetic disorders associated with genomic variation. *Genet Med* **18**, 1258–1268 (2016).

11. Richards, S. *et al.* Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* **17**, 405–24 (2015).

12. Gonsalves, S. G. *et al.* Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline for the Use of Potent Volatile Anesthetic Agents and Succinylcholine in the Context of RYR1 or CACNA1S Genotypes. *Clin. Pharmacol. Ther.* **105**, 1338–1344 (2019).

13. Yamaguchi, S., Brailey, L. L., Morizono, H., Bale, A. E. & Tuchman, M. Mutations and polymorphisms in the human ornithine transcarbamylase (OTC) gene. *Hum. Mutat.* **27**, 626–632 (2006).

14. Ghodke-Puranik, Y. *et al.* Valproic acid pathway: pharmacokinetics and pharmacodynamics. *Pharmacogenet. Genomics* **23**, 236–241 (2013).

15. Gonzalez, D. E. *et al.* Enzyme replacement therapy with velaglucerase alfa in Gaucher disease: Results from a randomized, double-blind, multinational, Phase 3 study. *Am. J. Hematol.* **88**, 166–171 (2013).

16. Hovelson, D. H. *et al.* Characterization of ADME gene variation in 21 populations by exome sequencing. *Pharmacogenet Genomics* **27**, 89–100 (2017).

17. Hopkins, P. M. *et al.* European Malignant Hyperthermia Group guidelines for investigation of malignant hyperthermia susceptibility. *BJA Br. J. Anaesth.* **115**, 531–539 (2015).

18. Johnston, J. J. *et al.* Variant curation expert panel recommendations for RYR1 pathogenicity classifications in malignant hyperthermia susceptibility. *Genet. Med.* **23**, 1288–1295 (2021).

19. Nezu, J. *et al.* Primary systemic carnitine deficiency is caused by mutations in a gene encoding sodium ion-dependent carnitine transporter. *Nat. Genet.* **21**, 91–94 (1999).

20. Toh, D. S. L., Yee, J. Y., Koo, S. H., Murray, M. & Lee, E. J. D. Genetic variations of the SLC22A5 gene in the Chinese and Indian populations of Singapore. *Drug Metab. Pharmacokinet.* **25**, 112–119 (2010).

21. Toh, D. S. L. *et al.* Functional analysis of pharmacogenetic variants of human organic cation/carnitine transporter 2 (hOCTN2) identified in Singaporean populations. *Biochem. Pharmacol.* **82**, 1692–1699 (2011).

22. Fischel-Ghodsian, N. Mitochondrial deafness mutations reviewed. *Hum Mutat* **13**, 261–70 (1999).

23. DiStefano, M. T., Hughes, M. Y., Patel, M. J., Wilcox, E. H. & Oza, A. M. Expert interpretation of genes and variants in hereditary hearing loss. *Med. Genet.* **32**, 109–115 (2020).

24. McDermott, J. H. *et al.* Clinical Pharmacogenetics Implementation Consortium (CPIC) guideline for the use of aminoglycosides based on MT-RNR1 genotype. *Clin. Pharmacol. Ther.* (2021).

25. Empey, P. E., Pratt, V. M., Hoffman, J. M., Caudle, K. E. & Klein, T. E. Expanding evidence leads to new pharmacogenomics payer coverage. *Genet. Med.* **23**, 830–832 (2021).